

P-2213

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-272686

(43)Date of publication of application : 08.10.1999

(51)Int.Cl.

G06F 17/30  
G06F 17/27

(21)Application number : 10-070688

(71)Applicant : NIPPON TELEGR & TELEPH CORP  
<NTT>

(22)Date of filing : 19.03.1998

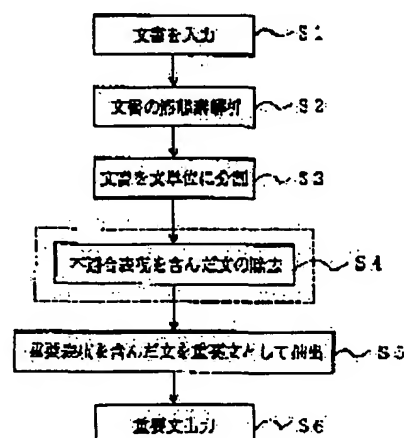
(72)Inventor : HORII MUNEYUKI  
MATSUOKA KOJI  
TAKAGI SHINICHIRO

(54) METHOD AND DEVICE FOR IMPORTANT DOCUMENT SENTENCE EXTRACTION AND RECORD MEDIUM WHERE IMPORTANT DOCUMENT SENTENCE EXTRACTING PROGRAM IS RECORDED

(57)Abstract:

PROBLEM TO BE SOLVED: To easily extract an important sentence from a document with high precision.

SOLUTION: An improper expression table wherein improper expressions are described and an important expression table wherein important expressions are described are prepared; and a morpheme analysis (S2) of an inputted document is taken, the analyzed document is divided (S3) into sentences, and sentences including improper expressions are removed (S4) from the document divided into the sentences by referring to the improper expression table. From the document from which the sentences including the improper expression have been removed, sentences including important sentences are extracted (S5) as important sentence by referring to the important expression table. Here, the process for removing the sentences including the improper expressions is omitted in some cases and the sentences including the important expression may be extracted as important sentence directly from the document divided into the sentences by referring to the important expression table.



## LEGAL STATUS

[Date of request for examination] 10.05.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

(18)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-272886

(43)公開日 平成11年(1999)10月8日

| (SI)Int.Cl. |       | 識別記号    |        | PI      |         |
|-------------|-------|---------|--------|---------|---------|
| G 0 6 F     | 17/30 | G 0 6 F | 15/401 | G 0 6 F | 3 2 0 A |
|             | 17/27 |         | 15/20  |         | 5 5 0 E |
|             |       |         | 15/38  |         | D       |
|             |       |         | 15/40  |         | 3 7 0 A |

審査請求 未請求 請求項の数 5 O L (全 7 頁)

|          |                  |         |   |
|----------|------------------|---------|---|
| (21)出願番号 | 特開平10-70688      | (71)出願人 | 000004228<br>日本電信電話株式会社<br>東京都千代田区大塚二丁目3番1号 |
| (22)公開日  | 平成10年(1998)3月19日 | (72)発明者 | 足井 毅之<br>東京都新宿区四新橋三丁目18番2号 日本<br>電信電話株式会社内  |
|          |                  | (72)発明者 | 松岡 浩司<br>東京都新宿区四新橋三丁目18番2号 日本<br>電信電話株式会社内  |
|          |                  | (72)発明者 | 高木 伸一郎<br>東京都新宿区四新橋三丁目18番2号 日本<br>電信電話株式会社内 |
|          |                  | (74)代理人 | 弁護士 鈴木 誠                                    |

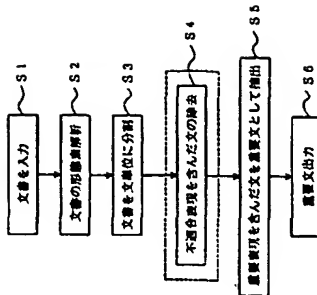
(54)【発明の名称】 文書重要文抽出方法、文書重要文抽出装置及び文書重要文抽出プログラムを記録した記録媒体

(57)【要約】

【課題】 文書中の重要文を、簡潔に且つ高い精度で抽出することを可能にする。

【解決手段】 重要文として不適切な表現を記述した不適切表現テールと、重要表現を記述した重要表現テールを用意し、入力された文書を形態素解析し (S2)、該形態素解析された文書を文単位に分割し (S3)、該文単位に分割された文書中から不適切表現テールを参照し不適切表現を含んだ文を取り除き (S4)、該不適切表現を含んだ文を取り除いた後の文書中から重要表現テールを参照して重要表現を含んだ文を重要文として抽出する (S5)。ここで、場合によっては、不適切表現を含んだ文を取り除く処理は省略し、文単位に分割された文書中から、重要表現テールを参照して直接重要表現を含んだ文を重要文として抽出することによい。

本発明の文書重要文抽出プログラムの処理フロー図





急対策会議を行いますので、ご出席をお願いいたします。

12 申し訳ありませんが、よろしくお願ひします。

【0017】不通文除却部23では、属文分割部22で分割された全ての文に対し、不通表現テーパーブル42に記述された表現のいずれかが含まれているかどうかを調査し、1つでも含まれていた場合、その文を、不通表現を含んだ文、すなわち重要文とはならない文として、属文分割部22から受け取った文書から取り除く。

【0018】図4は、不通表現テーパーブル42の例を示す。不通表現テーパーブル42は、文位置情報と形態素情報から構成される。文位置情報は、形態素情報で記述されている表現が含まれているかどうかをチェックする対象となる文位置を記述している。通常「\*」となっており、全ての文を対象とする。たとえば「先頭文」となっている文位置を記述している。形態素情報は、1組以上の接辞情報、形態素記号、形態素記号、品詞の4つ組から構成される。接辞情報は、次の形態素の位置を規定する。「next」は次の形態素が直後にあることを表し、「last」は次の形態素が直前にあることを表し、「next」は次の形態素が直後にあることを表し、「last」は次の形態素が直前にあることを表す。表現情報、形態素記号、品詞における「\*」は全てにマッチすることを表す。

【0019】上記分割された各文において、たとえば、

1 接続詞情報 高木様  
では、「様(文末)」の部分が不通表現テーパーブル42のテーパーブル1000の表現にマッチする。従って、文書から取り除かれる文となる。同様にして、  
2 三田商事の竹石です。  
は、「竹石です」の部分がテーパーブル1010にマッチする。

3 いつもお世話になっております。  
は、「世話になつ」の部分がテーパーブル10102にマッチする。

12 申し訳ありませんが、よろしくお願ひします。  
は、「が、よろしくお願ひし」の部分がテーパーブル10103にマッチする。なお、文番号12は、「」と「お願ひ」の間に「よろしく」が入っているが、テーパーブル10103の「」の接続情報は「ior」になっているので、「お願ひ」が属しているも構わない。

【0020】このようにして、不通文除却部23では、入力文書から文番号1、2、3、12の4文が取り除かれ、以下の文書が重要文抽出部23へ送出される。

5 貴社に納入して頂いた在庫管理システムが今朝ダウ

ンし、現在非常モードで稼働中です。

6 主な状況は以下の通りです。

7 ・端末からアクセスしても初期画面が表示されま

せん。

8 ・バックアップデータがリストアできません。

9 年末時期で業務に支障があり、早急に対処策を検討しなければなりません。

10 至急電話を頂けないでしょうか？

11 本日15時から17時に三田商事本社会議室で緊急対策会議を行いますので、ご出席をお願いいたします。

【0021】重要文抽出部24では、不通文除却部23から受け取った、不通表現を含んだ文の取り除かれた文書中の全ての文に対し、重要表現テーパーブル43に記述された表現のいずれかが含まれているかどうかを調査し、1つでも含まれていた場合、重要文として抽出する。

【0022】図5に、重要表現テーパーブル43の例を図4に示す。重要表現テーパーブル43の構成は、図4の不通表現テーパーブル42の構成と同じである。したがって、重要文抽出部24では不通文除却部23と同様の手法で照合を行う。

【0023】上記文例では、不通文除却部23から受け取った文書のうち、

4 システム緊急会議をお願いです。  
の「お願いです」の部分がテーパーブル10202にマッチする。

10 至急電話を頂けないでしょうか？  
の「か？」の部分がテーパーブル10200にマッチする。

11 本日15時から17時に三田商事本社会議室で緊急対策会議を行いますので、ご出席をお願いいたします。  
の「お願いし」の部分がテーパーブル10201にマッチする。なお、テーパーブル10204の形態素情報は、

9 年末時期で業務に支障があり、早急に対処策を検討しなければなりません。  
の「を」の部分とマッチするが、文位置情報は「先頭文」とあるので、先頭文でない文番号9は該当しない。

【0024】以上により、重要文抽出部24では、上記の3文、すなわち、

4 システム緊急会議をお願いです。

10 至急電話を頂けないでしょうか？

11 本日15時から17時に三田商事本社会議室で緊急対策会議を行いますので、ご出席をお願いいたします。

が重要文として抽出されることになる。

【0025】図2の実施例では、不通文除却部23で不通表現を含んだ文を取り除いた文書を対象として、重要文抽出部24で重要文を抽出しているが、不通文除却部23での処理を省略し、不通文を取り除かず、重要文抽出を行うことで、その場合、上記文例では、文番号1から12の文全てが重要文抽出部50への入力となるので、文番号4、10、11の3文以外に、

12 申し訳ありませんが、よろしくお願ひします。  
が抽出される。「(「お願いし」とテーパーブル10201がマッチ)。

【0026】したがって、抽出される重要文は、

4 システム緊急会議をお願いです。

10 至急電話を頂けないでしょうか？

11 本日15時から17時に三田商事本社会議室で緊急対策会議を行いますので、ご出席をお願いいたします。

12 申し訳ありませんが、よろしくお願ひします。の4文となる。

【0027】このように、不通文除却部23の処理を省略することで、それほど重要ではない文番号12も抽出されてしまうが、上記文例では、多少精度が落ちる程度で問題はない。

【0028】次に、他の具体例として、文書入力装置10で、以下の電子メール文書が入力された場合を考え

る。  
<入力>  
田中です。

4月25日に「第3回情報通信講習会」を開催します。  
今回のテーマは、「エージェンツ通信」です。佐藤さんのグループには関連の深いテーマだと思ひます。参加希望者がございましたら、田中までご連絡ください。多数の参加、お待ちしております。

【0029】形態素解析部21、属文分割部22を経て、

1 田中です。

2 4月25日に「第3回情報通信講習会」を開催します。

3 今回のテーマは、「エージェンツ通信」です。

4 佐藤さんのグループには関連の深いテーマだと思ひます。

5 参加希望者がございましたら、田中までご連絡ください。

【図1】本発明の文書重要文抽出方法の処理フロー図である。

【図2】本発明の文書重要文抽出装置の一実施例の概略ブロック図である。

【図3】形態素解析部23の内部構成の概略ブロック図である。

【図4】本発明に用いる不通文除却部23の内部構成の概略ブロック図である。

【図5】本発明に用いる重要表現テーパーブルの内部構成の概略ブロック図である。

【図6】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図7】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図8】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図9】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図10】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図11】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図12】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図13】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図14】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図15】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図16】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図17】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図18】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図19】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図20】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図21】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図22】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図23】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図24】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図25】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図26】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図27】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図28】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

【図29】本発明に用いる重要文抽出装置の一実施例の概略ブロック図である。

